REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From – To)
10-2003	Conference proceeding	Sep 2002 – Oct 2003

4. TITLE AND SUBTITLE

Improved target identification of correlated input data using recurrent neural networks and feature selection

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER
QAF185025203203 (AFOSR)
NNFMJH03682952 (ACC)

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)

Maj. Trevor I. Laine
Dr. Kenneth W. Bauer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology, Department of Operational Sciences
2950 Hobson Way, Bldg 640, Wright-Patterson AFB, OH 45433

8. PERFORMING ORGANIZATOIN REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOFR Attn: Major Juan Vasquez
801 N. Randolph St, Room 933, Arlington, VA 22203-1977
(703) 696-8431, e-mail: Juan.Vasquez@afosr.af.mil

ACC/DRS Attn: Chuck Sadowski
216 Hunting Ave, Room 106, Langley AFB, Va 23665-2777
(757) 764-1704, e-mail: Charles.Sadowski@langley.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR and ACC/DRSA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

For non-cooperative targets, combat ID may be accomplished by fusing data obtained from multiple sensors taken across time periods using ATR algorithms. With some ambiguity existing amongst fusion models, definitions are first developed to identify the specific type of fusion to be performed. Since input features extracted from sensor data for ATR algorithms are likely to contain significant correlation, models such as artificial neural networks that do not assume independent input data are a viable approach for fusion. An experiment was designed to assign generated temporal data with significant autocorrelation, crosscorrelation and noise into one of two classes. This feasibility study assesses use of an Elman recurrent neural network to perform fusion of multiple sensors with multiple looks to accomplish target identification. To improve classification accuracy, feature saliency screening was performed to select a subset of eight candidate input features with a signal-to-noise ratio and a network output sensitivity based measure. Both measures indicate a subset of about three of the original eight features should be retained. When comparing the two methods, both selection and ranking of salient features is consistent. Numerical results show the parsimonious subset of features improved generalization by significantly reducing the classification accuracy variance across multiple data sets and through time periods. Additionally, the reduced feature set yields an increase in the observed classification accuracy for the last time period of the external validation set.

15. SUBJECT TERMS
Combat Identification, Automatic Target Recognition, ATR, Sensor Fusion, Classification, Feature Saliency, Feature Selection, Neural Networks, Recurrent Networks, RNN

16. SECURITY CLASSIFCATION:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Trevor I Laine, AFIT/ENS
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (include area code)
U	U	U	UU	28	(937) 255-6565, x4350 e-mail: Trevor.Laine@afit.edu

IMPROVED TARGET IDENTIFICATION OF CORRELATED INPUT DATA USING RECURRENT NEURAL NETWORKS AND FEATURE SELECTION

Trevor I. Laine and Kenneth W. Bauer
Air Force Institute of Technology
Department of Operational Sciences
2950 Hobson Way, Bldg 640
Wright-Patterson AFB, OH 45433
Phone: (937) 255-6565 x4350 and x4328
FAX: (937) 656-4943
Email: Trevor.Laine@afit.edu, Kenneth.Bauer@afit.edu

71st MORS Symposium
WG 15 – Air Power and Combat Identification Analysis
12 June 2003

ABSTRACT

For non-cooperative targets, combat identification may be accomplished by fusing data obtained from multiple sensors taken across time periods using automatic target recognition (ATR) algorithms. With some ambiguity existing amongst fusion models, definitions are first developed to identify the specific type of fusion to be performed. Since input features extracted from sensor data for ATR algorithms are likely to contain significant levels of correlation, models such as artificial neural networks that do not assume independent input data are a viable approach for fusion. An experiment was designed to assign generated temporal data with significant autocorrelation, crosscorrelation and noise into one of two classes. This feasibility study assesses use of an Elman recurrent neural network to perform fusion of multiple sensors with multiple looks to accomplish target identification. To improve classification accuracy, feature saliency screening was performed to select a subset of eight candidate input features with a signal-to-noise ratio and a network output sensitivity based measure. Both measures indicate a subset of about three of the original eight features should be retained. When comparing the two methods, both selection and ranking of salient features is consistent. Numerical results show the parsimonious subset of features improved generalization by significantly reducing the classification accuracy variance across multiple data sets and through time periods. Additionally, the reduced feature set yields an increase in the observed classification accuracy for the last time period of the external validation set.

INDEX TERMS
Combat ID, Automatic Target Recognition, Sensor Fusion, Pattern Recognition, Neural Networks, Recurrent Neural Networks, Feature Saliency, Feature Selection.

Approved for public release; distribution unlimited

INTRODUCTION

With recent technological advancements in precision engagement and stealth, "if the enemy's key targets, target sets, or COGs (centers of gravity) can be found and *identified*, they are usually within airpower's reach" (Dept. of AF 2000: p42). Combat target identification (CID) is hence identified by *Air Force Doctrine Document (AFDD) 2-1: Air Warfare*, as one of the limiting factors in our ability to engage the enemy. An assessment of the current state of CID by Haspert (2000) concurs with this assessment of CID and goes on to state, "CID is often viewed as the weakest part of the military's kill chain." Where, the links in the complete kill chain may include: searching, detecting, tracking, classifying, identifying, assigning, fire control calculations, weapons launch, mid-course guidance, target acquisition by the weapon, terminal homing, fusing, target damage, and battle damage assessment.

While fusion is identified in *AFDD 2-5.2: Intelligence, Surveillance, and Reconnaissance Operations* (Dept. of AF 1999) as an AF principle to obtain high levels of confidence for combat target declarations, the details on fusion techniques for combat ID are not provided there or within the more specific guidance of *AFP 14-210: USAF Intelligence Targeting Guide* (Dept. of AF 1998). A review of open source literature identifies data fusion as a relatively new area for both DoD and non-DoD research, where improved estimates of unknown states can be obtained by combining information derived from multiple sources (Hall & Llinas 2001). As an emerging multidisciplinary field, Hall & Llinas (2001) state, "...there are disagreements in the data fusion community concerning which (fusion) method is best," and also emphasize each potential competing fusion technique should be considered and evaluated in the context of the specific task at hand.

To meet the USAF requirement of obtaining a minimum level of confidence before targets can be engaged, data from different sensors may be fused or if no class declaration has been made, data obtained from re-looks of an object through time may be fused. Thus, optimal methods are sought to fuse information as outlined above. This paper will first review common fusion taxonomies and develop applicable definitions for the specific use of fusion for combat ID. This research then demonstrates the feasible use of a Recurrent Neural Network (RNN) to perform fusion of sensor data. The selection of an optimal subset of salient sensor input features representative of the intrinsic dimensionality of the sensor data relevant for decision making is also assessed and compared for two competing feature saliency measures.

The remainder of this paper is organized as follows. First, a background of ATR is presented. Next, fusion definitions and hierarchical models are provided to facilitate a specific understanding of the fusion task at hand. Artificial neural networks (ANNs) and input feature selection are then introduced. The target identification experiment with input sensor/feature selection and conclusion follow.

AUTOMATIC TARGET RECOGNITION BACKGROUND

With combat ID identified as a current weakness in the military kill chain, methods to improve the ability of accurately identifying targets is highly desired. Technological advances in unmanned aerial vehicles for both combat and surveillance missions have also increased the desire to automatically engage a target without a man-in-the-loop. While a requirement for automatic target recognition (ATR) was identified in the 1960's, ATR development has continued through the 1990's and fully automatic target recognition has not yet been operationally fielded (Nasr 2003). Recent technological advances have overcome most data collection and processing requirements within time, space, and weight constraints; yet, the ability to accurately and consistently make target declarations across extended operational conditions (EOC) remains a challenge (Nasr 2003; Ross *et al.* 1999). To meet the challenge of performing ATR in EOCs, a feature space representation of Targets and Non-targets with maximum class separation is highly desired. Here the extracted features may be derived from physically independent sensors, similar sensors on different platforms, or from the same sensor through time.

A general representation of the ATD/R process from Schroeder (2002) is included as Figure 1 and can be used to help visualize the steps required for ATR, where each forward step looks to refine the assessment of an object observed after detection.

Detect → Discriminate → Classify → Recognize → Identify

Figure 1. Process model of Automatic Target Detection/Recognition (ATD/R) as presented by Schroeder (2002) and applied to SAR imagery.

The process blocks are defined as:

- ***Detect***: Identify a Region of Interest (ROI) for analysis of a potential target
- ***Discriminate***: Binary decision-target either present or not present in ROI
- ***Classify***: Targets grouped into general class, e.g. Tank, Armored Personnel Carrier
- ***Recognize***: Subdivision of class types, e.g. T-72 tank
- ***Identify***: Unique identification of a target, e.g. assignment of serial number

Progressing from *Detect* to *Identify*, increased levels of data resolution or data from multiple looks or sources may be required to further refine the assessment of a potential target. To proceed past the *Discriminate* block, fusion of data from multiple sources may be required to meet predetermined confidence levels to either declare a ROI as having a hostile target or not having one.

To meet requirements across many operational environments, data from a single sensor type, even if acquired from multiple looks, is likely to be inadequate. Table 1

shows basic sensor properties for mature sensors typically considered for AF combat ID applications. Sensor types include forward looking infrared (FLIR), synthetic aperture radar (SAR) and optical sensors in the visible spectrum. Because of the ability for FLIR and SAR data to be collected day or night, and because FLIR and SAR complement each other well, much of the ATR sensor fusion research has focused on fusion of these two sensor types (Nasr, 2003).

Table 1. Sensor characteristics as identified by Nasr (2003) where an "X" indicates the specific sensor properties, environmental conditions where the sensors performs well, and counter-measures each sensor can effectively defeat.

Sensor Type	**FLIR**	**SAR**	**Visible**
Sensor Properties			
Active		X	
Passive	X		X
Rapid Scan	X		X
Environmental Conditions			
Day & Night	X	X	
Adverse Weather		X	
Smoke & Dust		X	
Clutter			
Counter-measures			
Corner Reflectors	X		X
Camouflage	X		
Decoys/Flare		X	X

Future systems may include newer multispectral imagery (MSI) and hyperspectral imagery (HSI), which collect data from visible through thermal IR frequencies within numerous frequency bands all from a single sensor. While significant ATR improvements were obtained by Young *et al.* (2001) by fusing SAR and MSI data, only 5 of the 12 candidate MSI bands were used. These 5 bands represented the largest spectral separation of the MSI sensor with the greatest chance of independent information. With hundreds of spectral bands, HSI imagery will provide increased spectral resolution of targets, but this increase may likely come with a lower Signal-to-Noise Ratio (SNR) as less reflected or emitted energy is available for sensor collection across the smaller spectral frequency bands (Landgrebe 2002). Because of high correlation levels between neighboring spectral bands, HSI data collected for numerous frequency bands may be no better for classification problems than MSI data (Gat *et al.* 1997). Also, collection of HSI data often produces very sparse data that can be projected into lower dimensions with minimal loss of information (Jimenez & Landgrebe 1998). Research to determine the optimal frequency bands may employ input feature saliency screening techniques in attempt to determine the underlying dimensionality of those features providing for best

class separation. Further, selection of optimal parsimonious salient features from high-dimensional spectral data may help to determine an optimal number of frequency bands and an optimal mix of sensors used to collect sets of best features.

While some features derived from passive visual or thermal sensors and reflected radar energy each containing different noise sources may be statistically independent, multiple looks by a single sensor across the time continuum are likely to contain significant correlation. If a fusion algorithm assumes independent input data for real-time ATR, violation of this assumption may overestimate performance when significant correlation is present. As stated by Dudgeon (1998: p22):

> The assumption of independence is often justified, but in some cases it is not, and it may lead to inaccurate estimates of performance. Generally, independence between two random variables can be used as the limiting case where the value of one variable has no correlation with and conveys no information about the value of the other.

Despite the correlation problem, many CID applications may require additional information to increase confidence after a "no class designation" label assignment. As the only source available for additional target information, re-looks by a sensor in close temporal proximity may be obtained. These multiple looks are hypothesized to have high levels of positive correlation and may provide relatively little new information about the object. Literature from the radar community (Chitroub *et al.* 2002; Costantini *et al.*; 1997, Lee *et al.* 1994) indicate high correlation levels are indeed expected between SAR imagery data obtained from continuous re-looks of an area. Current image processing approaches use these multiple correlated looks to refine a single image by reducing noise in the image as additional SAR images are obtained. While this SAR imagery refinement is primarily done for visual interpretation and methods are not presented to make subsequent object class declarations, it does suggest a basic analysis framework for temporally collected sensor data. Similar to an image refinement type process, as temporal information is gathered, ATR may benefit from algorithms designed to update and refine class estimates based on obtaining new, albeit correlated, information.

To further complicate CID, growth in the total volume of information available in the current "information age" and the resulting dimensionality of data available for fusion continues to grow. Sources of data growth include increases in sensor resolution, bandwidth to share information, the number of ISR platforms (e.g. UAVs and satellites), and new sensor types like MSI and HSI, which generate tens or hundreds of values for each pixel. Moreover, fusing multilook sensor data increases the dimensionality of the fusion process, and temporal fusion is not well understood (Dasarathy 1997: p27). Some current fusion research looks to understand the effects of input data growth, where pattern recognition or target ID is dominated by methodologies using a feature vector derived from sensor data to represent each object in a feature space with defined class boundaries (Hall & Llinas 1997: p19-20). While many techniques for pattern recognition using feature vector input are available to the analyst, inclusive of neural networks approaches,

Hall & Llinas (1997) note high quality input data may be more important than the specific classification model selected for use, where:

> ...the ultimate success of these methods depends upon the ability to select good features. (Good features are those which provide excellent class separability in feature space, while bad features are those which result in greatly overlapping areas in feature space for several classes of targets.)

They go on to state, "...more research is needed to guide the selection of features and to incorporate explicit knowledge about target classes," (such as other intelligence information). Guidance for the selection of features is offered by Looney (1997: ch10) with two goals of mapping classification data into a feature space summarized as:

1. Retain as much relevant information as possible
2. Remove as much redundant information and extraneous noise as possible

In one approach to this problem, the estimated linear correlation is typically used to measure the degree of association and linear dependence between any two random variables or features. This correlation is an efficient measure to indicate possible redundancy or dependence between features. Yet, with non-linear classifiers, including ANN models, this measure of linear correlation may not be sufficient to screen candidate input features. Classifier specific feature selection approaches may be preferred. This research will use recurrent ANN models with feature selection algorithms in attempt to find a subset of input features to increase discrimination between classes. With both significant noise and correlation inherent to the candidate input features, a subset of salient features will be compared to the complete set of candidate input features.

FUSION PROCESS MODELS, TAXONOMY AND DEFINITIONS

The following section provides a brief review of common fusion process models and introduces fusion level definitions. Because there does not seem to be universally accepted definitions for such terms as sensor fusion and classifier fusion, an attempt was made to develop definitions of sensor fusion levels based on physical characteristics of both the input data and use of the model output. The process models provide current nomenclature and definitions from various fusion communities and serve as a foundation to develop less ambiguous definitions to characterize levels of sensor fusion. Models reviewed include the JDL Model (Hall & Llinas 2001; Steinberg *et al.* 1999; Hall & Llinas 1997; Waltz & Llinas 1990), UK Intelligence Cycle Model (Bedworth 1999), Boyd OODA Loop (Boyd 1987), Waterfall Model (Bedworth 1999), Omnibus Model (Bedworth & O'Brien 2000) and the Dasarathy I/O Fusion Model (Dasarathy 1997). From these models, all but the Dasarathy I/O model characterize fusion levels by the tasks or functional use of the output data.

Prominent within fusion literature, the JDL model was first proposed by the Data Fusion Working Group chartered to study information fusion for DoD applications. This working group was established in 1986 and subsequently created the JDL model and a Data Fusion Lexicon (Hall & Llinas, 1997: p11). With an original emphasis on tactical targeting issues, the initial model was developed for military specific applications, but was later revised to encompass growing nonmilitary applications such as manufacturing processes, complex system monitoring, robotics, and medical applications. Revisions to the to the JDL data fusion model are presented in (Steinberg *et al.* 1999) where fusion levels are a categorization of output data functions. The revised JDL model is presented in Figure 2, where the data fusion domain includes Levels 0-4 and Database Management. Various sources of local input data have also been included for illustrative purposes corresponding to a CID application.

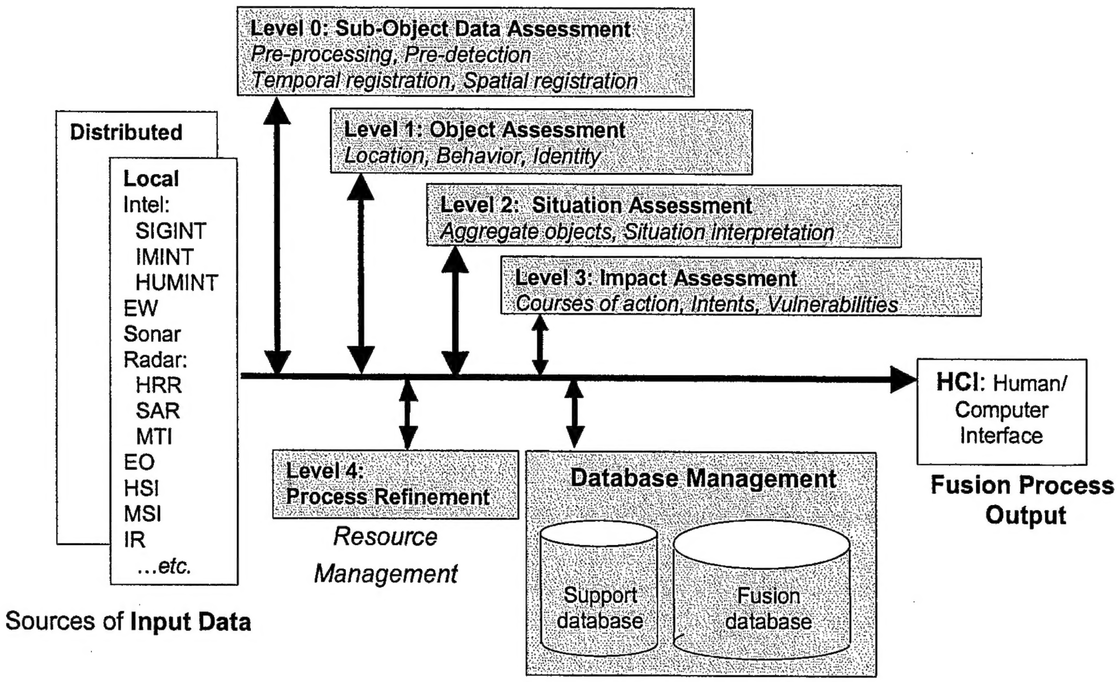

Figure 2. The revised JDL Data Fusion Model.

A recent fusion model is the Omnibus model (Bedworth & O'Brien 2000) that encompasses cyclic UK intelligence and Boyd OODA loop properties. This Omnibus model incorporates the finer definitions from the Waterfall model and can be mapped to both the JDL model based on tasks and to the Dasarathy model based on the input/output characteristics of the fusion occurring within any of the four Omnibus model levels of fusion: sensor data, feature, soft decision, and hard decision. To note, feature level fusion is included within the Orient process, with the selection of correct features for pattern

recognition (target identification) identified as one of the current limitations of feature fusion (Bedworth & O'Brien 2000).

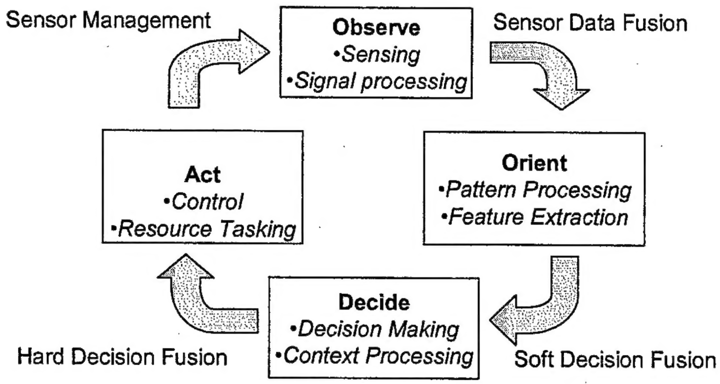

Figure 3. Omnibus model for data fusion (Bedworth & O'Brien, 2000).

In contrast to the previous models, the Dasarathy fusion model identifies levels of fusion based on the type of input information being fused and the characteristics of the output data, resulting in an I/O-based characterization model (Dasarathy 1997). The three types of input and output include:

- ***Decisions***: Belief values
- ***Features***: Intermediate level values
- ***Data***: Observed raw data with minimal manipulation

These input and output labels lead to five distinct types of fusion, identified in Table 2. An illustration of the various types of I/O fusion is presented in Figure 4, where fusion can occur on parallel or upward arrows and may occur repeatedly within a system.

Table 2. The five levels of information fusion in the Dasarathy model.

Input	Output	Notation	Description/Analogy
Data	Data	DAI-DAO	Data-level fusion
Data	Features	DAI-FEO	Feature selection; Features extraction
Features	Features	FEI-FEO	Feature-level fusion
Features	Decisions	FEI-DEO	Pattern recognition; Pattern processing
Decisions	Decisions	DEI-DEO	Decision-level fusion

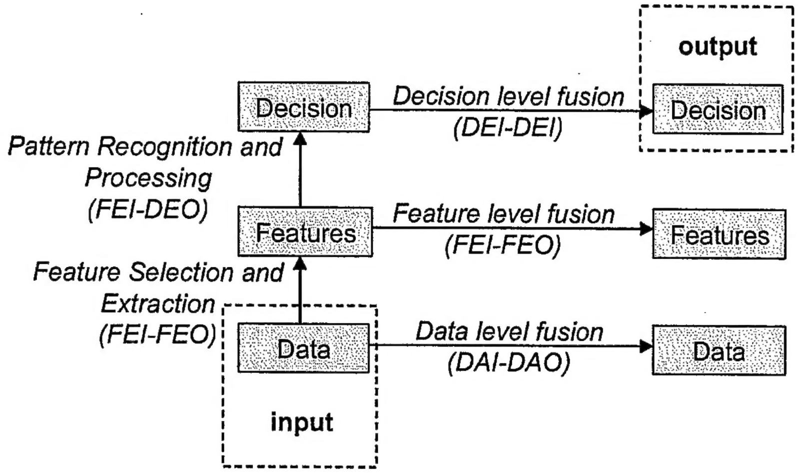

Figure 4. Dasarathy I/O fusion model, as derived from (Dasarathy 1997)

A summary and comparison for each of the common models is presented in Table 3. This table builds upon and modifies a similar table presented by Bedworth and O'Brien (2000). Specific differences include minor changes in how the fusion levels are mapped to activities, modifications of activity titles, and inclusion of the Omnibus and Dasarathy I/O models.

Table 3. Comparison of fusion model components as a function of activity performed.

<table>
<tr><th>Activity</th><th>UK Intelligence Cycle</th><th>Boyd OODA Loop</th><th>Revised JDL model</th><th>Waterfall model</th><th>Dasarathy model</th><th>Omnibus model</th></tr>
<tr><td>Action</td><td rowspan="2">Disseminate</td><td>Act</td><td>HCI</td><td></td><td></td><td>Act</td></tr>
<tr><td>Decision making</td><td>Decide</td><td>Level 4</td><td>Decision making</td><td rowspan="2">DEI-DEO
FEI-DEO</td><td rowspan="2">Decide</td></tr>
<tr><td>Impact assessment</td><td rowspan="2">Evaluate</td><td rowspan="3">Orient</td><td>Level 3</td><td rowspan="2">Situation assessment</td></tr>
<tr><td>Situation assessment</td><td>Level 2</td><td rowspan="2">FEI-DEO
FEI-FEO
DAI-FEO</td><td rowspan="2">Orient</td></tr>
<tr><td>Information processing</td><td rowspan="2">Collate</td><td>Level 1</td><td>Pattern processing
Feature extraction</td></tr>
<tr><td>Data processing</td><td rowspan="2">Observe</td><td>Level 0</td><td>Signal processing</td><td rowspan="2">DAI-FEO
DAI-DAO</td><td rowspan="2">Observe</td></tr>
<tr><td>Detection</td><td>Collect</td><td>Input</td><td>Sensing</td></tr>
</table>

Fusion Level Definitions

Since some ambiguity exists between defined fusion levels for models based on decision processes, more precise definitions are desired. Less ambiguity exists when classifying fusion levels according to the transformation of the information during a fusion process. Since appropriate quantitative fusion techniques can be associated with a particular type of input data and desired transformation, definitions of fusion levels incorporating the Dasarathy I/O (Dasarathy 1997) characterization are adopted. The application of the fusion output will then dictate an appropriate mapping to each of the fusion models as illustrated in Table 3. Definitions for five levels of information fusion are as follow:

- ***Data Level Fusion*:** Fusion performed at the lowest level, combining raw data, registration information, and possible noise reduction. Includes DAI-DAO fusion along with aspects of data preprocessing and Level 0 fusion from the JDL model.

- ***Feature Level Fusion*:** Fusion performed to generate a new representation of the data by mapping it into a feature space. Includes both DAI-FEO and FEI-FEO fusion processes, commonly included in as JDL Level 1 fusion or object refinement.

- ***Identity Level Fusion*:** Fusion of features to generate a *posterior* estimate of class membership or other state used to quantify an object. Includes the FEI-DEO process, typified as pattern recognition and included as a distinct part of JDL Level 1 fusion.

- ***Decision Level 1 Fusion*:** Fusion of object assessments leading to a refinement in the current *posterior* estimated state of the object of interest. Includes DEI-DEO fusion, which may further refine the object assessment and is included in JDL Level 1.

- ***Decision Level 2 Fusion*:** Fusion of object Decision Level assessments or possibly different object Features, leading to a refinement in the current estimated situational state. Includes DEI-DEO fusion processes and possibly FEI-DEO processes. This matches the JDL model nomenclature where object aggregation first occurs at Level 2.

To perform fusion at the various levels, numerous quantitative techniques are available and are well documented within the literature (e.g. see Hall & Llinas 2001). Neural networks are one modeling technique used for Identity and Decision Level 1 Fusion (Hall & Llinas 1997). Other techniques successfully employed for Identity and Decision Fusion to estimate an object's identity include various pattern recognition methods: cluster algorithms, template methods, statistical methods and probabilistic methods such as those presented in (Ralston 1999) and (Haspert 2000). From the definitions presented above, ANNs and RNNs can be applied to perform fusion at the Feature, Identity, Decision Level 1, or at any combination of these Levels. A primary research objective is use of one big net (OBN) as a fusion tool, to determine an optimal class estimate or label for a single object, where features derived from multiple sensors may be fused or estimates or labels from different sensors may be fused together. While the input feature selection experiment within this paper represents Identity Level Fusion, the feature selection techniques may be applied for fusion of features derived from the same or different

sensors, posterior estimates from different ATR systems or from class labels provided from different sensors or other intelligence sources.

INTRODUCTION TO ARTIFICIAL NEURAL NETWORKS & FEATURE SELECTION

A current ATR challenge, i.e. fusion of sensor data from different sources through time to make Target or Non-target declarations, being addressed by this research is presented in Figure 5. Within this ATR application, multiple re-looks can be performed to obtain additional sensor information for an object of interest prior to making a final class declaration. An appropriate classification model, capable of effectively modeling temporal data with potentially high levels of correlation from one look to the next, is desired.

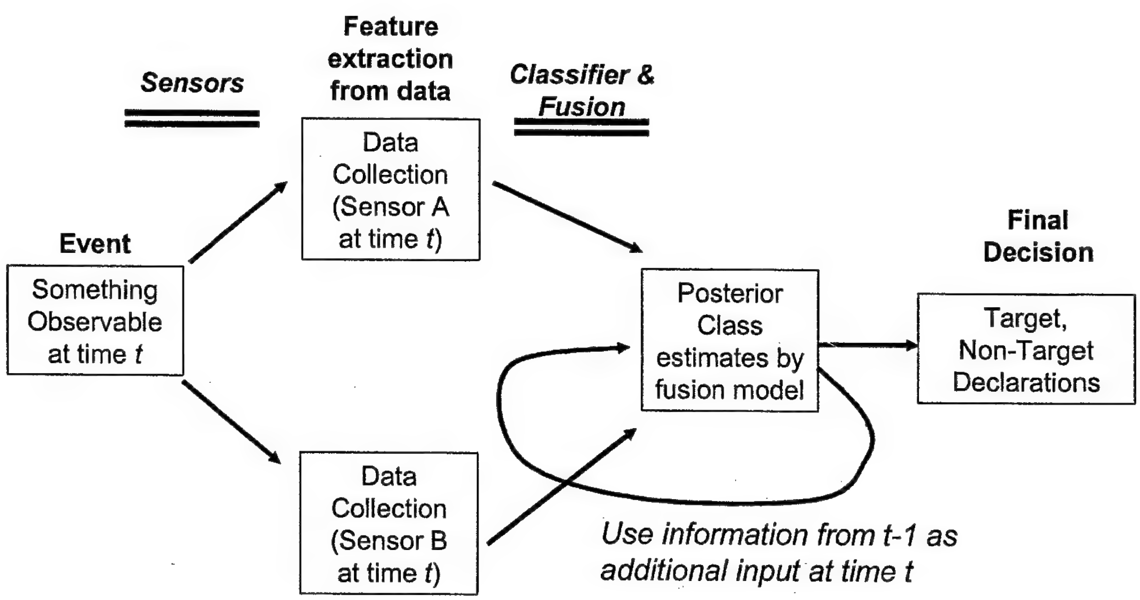

Figure 5. Sensor fusion process model representative of Decision Level 1 fusion where posterior class estimates may be refined as additional sensor data is obtained through time.

To perform this fusion research, neural network models are used for several reasons. Figure 6 represents a fully connected multilayer perceptron (MLP) ANN. While often viewed as a black box, these models are theoretically capable to perform any mathematical mapping from an input to output space with any desired degree of accuracy provided the number of hidden nodes is sufficiently large enough (Hornik *et al.* 1989, 1990). MLP ANNs offer a nonparametric approach to generate a mapping for input data with no assumed distribution or independence requirement between variables, to a desired output space. In addition, ANNs learn and may even adapt to new training data to obtain optimal parameter settings. Some drawbacks of ANNs include the expense of an available training data set fully representative of desired input and output spaces, along with the computational complexity of the training process, and a lack of decision insight. Yet, because they do not require assumptions of the input data structure, they are fully capable of integrating sensor features, estimated class probabilities and binary class labels, each of which may contain significant correlation between and across features; thus, ANNs allow for flexible sensor fusion via a one big net model.

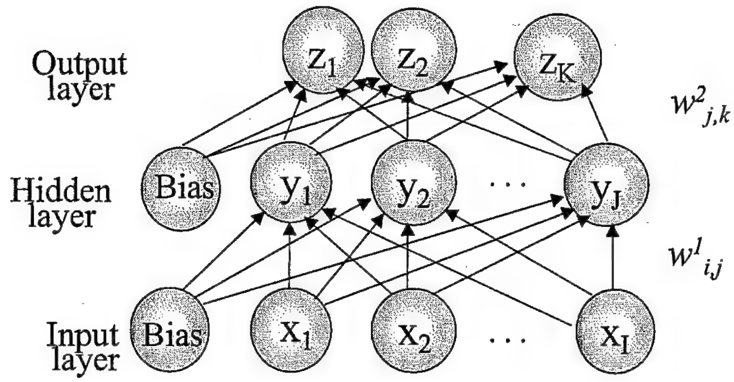

Figure 6. Multilayer Perceptron (MLP) Artificial Neural Network (ANN).

The output from such a MLP ANN for the *n*th input vector (z^n) can be computed as:

$$k^{\text{th}} \text{ neural network output} = z_k^n = f\left(\sum_{j=1}^{J} w^2{}_{j,k} x_j^1\right) \qquad (1)$$

where,

- J is the number of hidden nodes
- $f(\mathrm{a}) = 1/(1 + e^{-\mathrm{a}})$ is a typical sigmoidal activation function
- $w^2_{j,k}$ is the weight from hidden node j to output node k
- x_0^1 is the hidden layer bias term and is set equal to 1
- $x_j^1 = f(\Sigma\, w^1_{i,j} x_i^n)$ is the output of hidden node j and is summed from $i = 1$ to M
- M is the number of input features
- $w^1_{i,j}$ is the weight from input node i to hidden node j
- x_0^n is the input layer bias term and is set equal to 1
- x_i^n is the i^{th} input feature of the n^{th} input vector

While an ANN with proper architecture has been proven capable of universal function approximation, it may only explicitly model temporal relations in static time. Since a strong temporal component may be hypothesized for many pattern recognition applications such as financial forecasting or target tracking and identification, an ANN model is desired that allows for the implicit encoding of time. The Elman RNN includes internal feedback and the ability to model temporal patterns (Elman 1990). With an architecture similar to ANNs, an Elman RNN adds internal feedback to the model with each hidden layer output from time *t* included as input model at time *t+1*. Figure 7 shows an Elman RNN, with I input features, J context nodes, J hidden nodes and K outputs, where feedback is accomplished by the context nodes in Figure 7. Similar to ANNs, Elman RNN hidden and output layer perceptrons have associated activation functions, typically nonlinear sigmoid, hyperbolic tangent, or linear depending on the application. The hidden layer output included as context node input for the next data observation facilitate a dynamic memory for temporal patterns. By having internal feedback, the Elman RNN implicitly

models temporal patterns (Elman 1990) and has been proven to have the computational power of any finite state machine given a sufficiently large enough architecture (Giles & Omlin 2001; Kremer 1995). Further, the Elman RNN has increased modeling flexibility over another common RNN, the Jordan RNN, which uses the final network output from time t as context node input at time $t+1$ (Calvert & Kremer 2001).

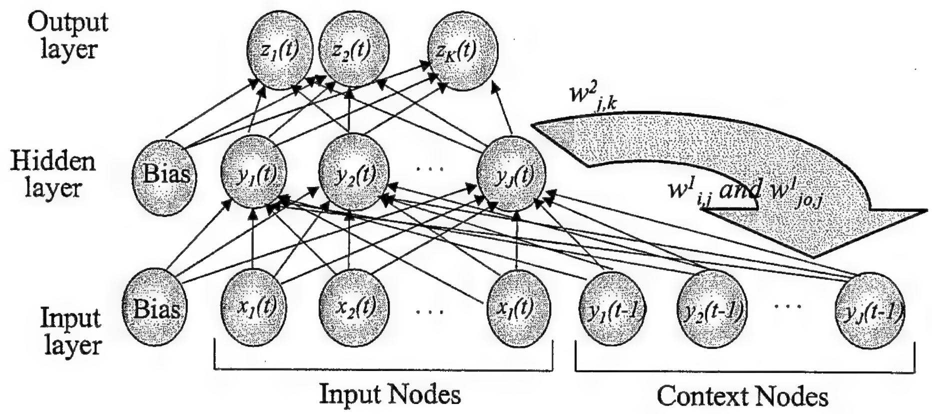

Figure 7. Elman Recurrent Neural Network (RNN).

Feature Selection for Pattern Recognition and Neural Networks

While properly configured neural networks can approximate any function, they are dependent on the quality of input data from which they learn or adjust their weight parameters. For statistical pattern recognition applications, it is well documented that too many features may decrease classification performance, since the number of observations must grow exponentially as the number of features increases to maintain the same sampling density. This "curse of dimensionality" (Bishop 1995) phenomenon, suggests feature reduction should be performed to improve results when limited data observations with sparse, high-dimensional input data are collected (Jackson & Landgrebe 2001). This section will focus on the comparison and assessment of two different input feature screening techniques to improve classification accuracy for a RNN. This research was initially presented in (Laine & Bauer 2003) and demonstrates use of an Elman RNN for Identity Level Fusion of temporal target patterns, where a subset of salient features is selected from candidate input variables known to contain significant correlation and noise.

In order to improve the model's accuracy, a reduced feature set representative of the underlying salient input feature space is desired. Feature engineering includes the extraction of salient features by finding a mapping to project P-dimensional input data onto M-dimensional space where $M < P$. Current literature identified few methodologies for RNN feature selection, with feature selection defined as a special case of feature extraction whereby the M-dimensional space corresponds to a subset of P collected potential input features. Research by Greene (1998), Greene *et al.* (1997, 2000), Utans *et al.* (1995) and Moody (1998) use RNN saliency metrics based on model weights and output error associated with input features. Since limited RNN saliency methods were identified, a broader review was undertaken of recent ANN feature

selection techniques that may be applicable to RNNs. Similar to the methods of Greene *et al.* and Moody *et al.*, other recent research is divided between techniques using ANN model weights (Castellano & Fanelli 2000; Lazzerini & Marcelloni 2002; Mak & Blanning) or model output (Feraud & Clerot 2002; Kwak & Choi 2002; Piramuthu 1999; Verikas & Bacauskiene 2002; Zhang & Sun 2002) with entropy measures associated with model output used by Piramuthu (1999) and Verikas & Bacauskiene (2002) and a tabu search based on observed model output employed by Zhang & Sun (2002).

The two feature screening techniques selected for comparison with an Elman RNN in this experiment are the Signal-to-Noise Ratio (SNR) feature screening introduced for ANN use by Bauer *et al.* (2000) and first applied to an Elman RNN by Greene (1998) and Sensitivity Based Pruning (SBP) as developed by Moody and presented in (Moody 1998; Utans *et al.* 1995) for general neural network use. These methods represent proven network parameter and output based saliency measures that will now be applied and compared using a RNN.

The SNR saliency measure is computed using the first layer weights of a trained RNN as

$$SNR_i = 10 \cdot \log_{base10}\left(\frac{\sum_{j=1}^{J}\left(w_{i,j}^{1}\right)^2}{\sum_{j=1}^{J}\left(w_{N,j}^{1}\right)^2}\right) \tag{1}$$

where SNR_i is the value of the SNR saliency measure for feature i, J is the number of hidden nodes, $w_{i,j}^{1}$ is the first layer weight from input node i to hidden node j, and $w_{N,j}^{1}$ is the first layer weight from an injected noise input node N to hidden node j. All feature inputs, including the randomly generated noise, are normalized. The scaled logarithmic transformation of the ratio converts the saliency measure to a decibel scale. The idea behind the SNR saliency measure is relevant features will have a SNR_i significantly greater than 0, while noise-like features will have a SNR_i saliency value close to or less than 0. The SNR saliency measure provides a way to rank order features from most relevant to least relevant and has been shown to be is statistically equivalent (Greene 1998) to that of Ruck's partial derivative based saliency measure (Ruck *et al.* 1990) and Tarr's weight based saliency measure (Tarr 1991) for ANNs. In addition, SNR feature selection has been successfully employed for fusion of correlated features derived from multiple sensors (Laine *et al.* 2002; Greene 1998) with an ANN, and feasibility has been demonstrated for time delayed neural nets (TDNNs) and RNNs by Greene (1998).

Like the SNR saliency measure, Sensitivity Based Pruning (SBP) associates a saliency measure to each input feature. The sensitivity measure S_i for each of i features is calculated by assessing the effect of replacing each input feature with the mean value of that feature (Moody 1998; Utans *et al.* 1995) and can be calculated once a RNN is trained as

$$S_i = MSE\left(\overline{x}_i\right) - MSE\left(x_{ip}\right) \tag{2}$$

where $MSE(x_{ip})$ is the mean square error of the RNN for all p exemplars and $MSE(\overline{x}_i)$ is the MSE when an average value is assigned to input feature i. If using the average value of a feature for all exemplars increases the MSE, S_i will be positive and considered salient, and the feature associated with the largest value of S_i is deemed the most salient feature. Thus, S_i values can be

used to rank order the relative saliency of input features for any ANN. If the input features have been normalized with a mean of zero prior to training the network, S_i can be computed simply by evaluating the trained RNN and setting each input feature to 0. Applications of SBP by Moody *et al.* for continuous financial time series prediction compute S_i for the training data, iteratively train and remove a feature from the network, then seek to select a parsimonious subset of features that minimizes prediction risk of a forecast. For this pattern classification experiment with 2 target types, the goal is to find a reduced feature set that maximizes classification accuracy (CA) and generalizes well to independent validation data. To compare the SBP and SNR measures, the SBP metric is implemented similar to the SNR measure, with S_i calculated from the training-test set to provide a measure of the RNNs ability to generalize well to new patterns and CA is calculated as

$$CA = \frac{\text{Number exemplars classified correctly.}}{\text{Total number of exemplars}} \tag{3}$$

Therefore, instead of prediction risk, CA is used to determine a final set of parsimonious salient features to retain for effective discrimination between two target classes.

Input Feature Saliency Screening

SNR and SBP screening methods use the saliency metrics from the previous section to obtain parsimonious sets of salient features while retaining good classification accuracy as features are removed backward stepwise. The experiment was performed using *Matlab 6.1* with the *Neural Network Toolbox*. RNNs were initialized with 8 hidden nodes and 2 output nodes with hyperbolic tangent and sigmoid transfer functions respectively. The desired outputs were set to 0.9 and 0.1 for correct and incorrect classes, and the classification decision was assigned based on the maximum of the two observed outputs. All networks were trained using gradient descent with momentum and an adaptive learning rate for a maximum of 1000 epochs. Most training stopped early after the training-test set MSE failed to improve after 200 epochs. The RNN weights associated with the minimum training-test set MSE were retained to be used as the trained network. Following are the steps to determine reduced salient feature sets:

1. Introduce a Uniform (0,1) noise feature, x_N, to the initial features (for SNR only).
2. Preprocess all features with mean zero and unit variance.
3. Initialize the RNN weights via the Nguyen & Widrow (1990) method.
4. Initialize input layer weights as uniform [-0.01, 0.01] (for SNR only).
5. Train the RNN and retain the weights that minimize the MSE of the test set.
6. Identify the least salient feature with the lowest SNR_i or S_i saliency metric.
7. Remove the least salient feature from the RNN.
8. Repeat steps 5, 6, and 7 until all features in the initial set have been removed.
9. Plot the training-test set classification accuracy (CA) as individual features are removed.
10. Retain the first feature whose removal caused a significant decrease in the training-test set CA, as well as all features removed after the first salient feature was identified.

Both screening methods seek to find a parsimonious set of input features representative of the underlying input feature space dimensionality. This is accomplished by reducing the features

used to discriminate between classes, such as removing one of two highly correlated input features. In previous research the SNR screening method has produced a reduced number of input features for an ANN while maintaining or improving classification accuracy for independent validation sets (Bauer *et al.* 2000; Greene *et al.* 2000; Laine *et al.* 2002).

INPUT FEATURE SALIENCY EXPERIMENT

To assess the utility and compare differences of a weight based and a performance based saliency measure in an Elman RNN, an experiment was designed with generated data. The generated data was inspired from observed data for 2 geosynchronous satellite types, processed by a Johnson filter and observed through time. This real data included the magnitude, corrected for distance, in disjoint red and blue electro-optical frequency bands, with temporal trends associated with the rotation of the earth, reflection from the sun and other atmospheric effects. A total of 8 input features were generated and each lasted for 10 time units. Data sets were comprised of 10 random sequences of each satellite type, producing 200 total observations in each data set. Three features were generated from a known parabolic "red" signature corrupted with 3 varying degrees of white noise. Similarly, an additional 3 features were generated from a decreasing logarithmic "blue" signature. The color types represented features derived from disjoint portions of the visible spectrum. An example of the generated data with the lowest levels of white noise is included as Figure 8, where no feature provided for linear separation of classes.

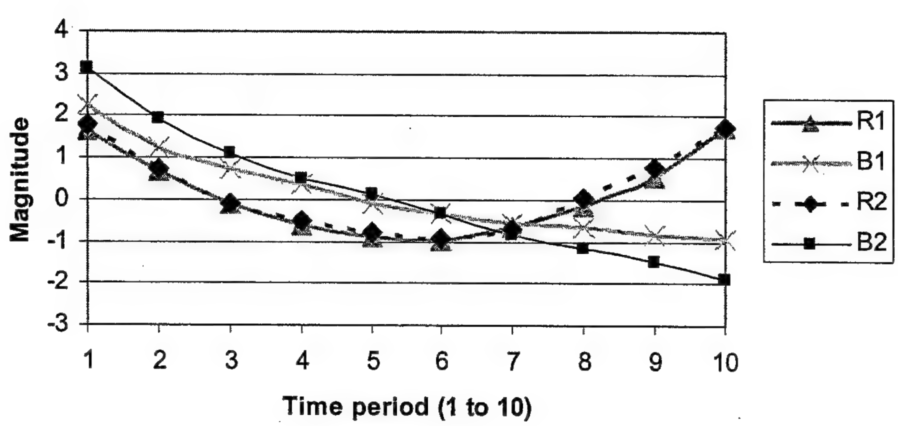

Figure 8. Normalized "red" (parabolic pattern) and "blue" (nonlinear decreasing) data with lowest noise corruption for target type 1 (R1 & B1) and target type 2 (R2 & B2).

The 6 "red" and "blue" features can be interpreted as being derived from 3 individual sensors each capable of obtaining data in the appropriate spectral bands, where each sensor may have different noise levels associated with sensor resolution or other feature extraction processes. Two additional features were constructed with significantly higher levels of noise and no significant correlation to the 2 classes, which may represent a sensor with little value to the identification task-at-hand. Since the data were generated from continuous functions of time with varying magnitudes of random noise added, autocorrelation was statistically significant and

crosscorrelation between variables derived from the same "color" was also statistically significant, as may be expected by sensors with different resolution observing the same target. Figure 9 is provided to illustrate such expected levels of correlation within one generated data set. A total of 20 data sets were generated for use as training, training-test, and validation sets. Training data was used to calculate error and update network weights, the training-test set was used to assess the trained RNN and stop training before over-fitting occurred, and the validation set was used to assess the final RNN on data not used for training the network weight parameters.

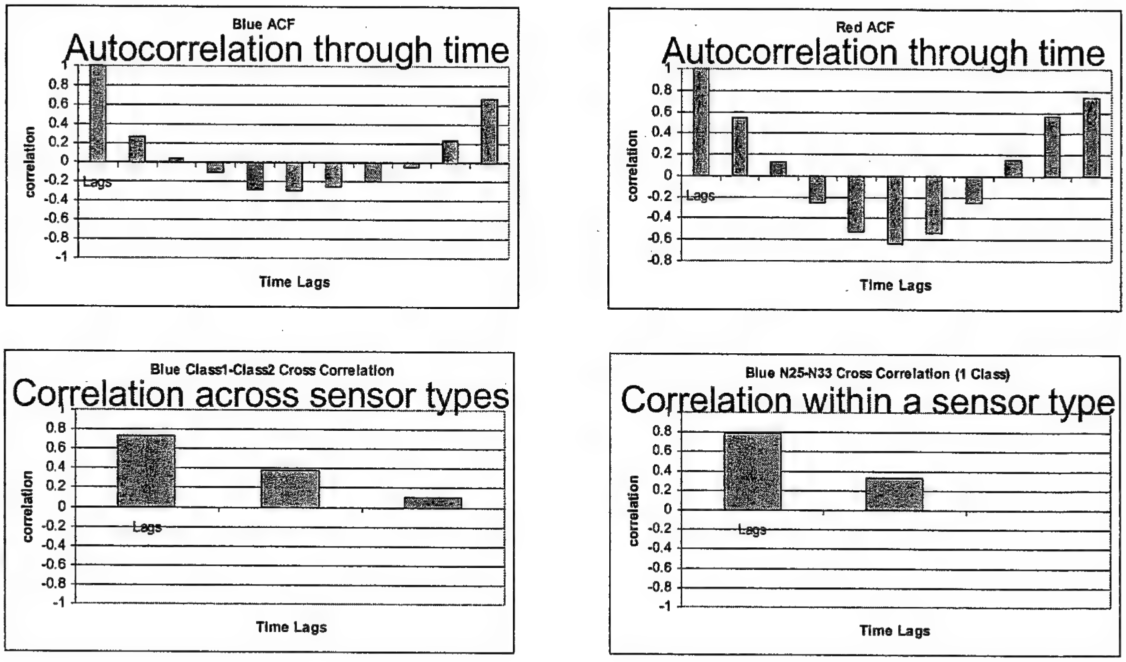

Figure 9. Representative sample of correlation estimates for 1 of 20 generated data sets. Note: correlation magnitudes > 0.20 are statistically significant.

Feature Saliency Screening Results

A sample graph of a SNR feature screening run as performed using an Elman RNN for the 2-class satellite experiment is presented as Figure 10. Similar plots were produced for the SBP screening. For both methods, the maximum CA obtained for the test set is used to select the parsimonious set of input features. As features are removed, the training set CA decreases slowly while the test set CA has an increasing trend as 0 to 5 features are removed, providing evidence of a viable reduced feature set. With 6 features removed, both CA values decrease signifying a possible loss in salient information contained by the input features used for class assignment. Thus, Figure 10 recommends a parsimonious set of 3 salient features (obtained with 5 features removed). Of interest is the convergence of the training and test set CA as features are removed, potentially indicative of improved generalization.

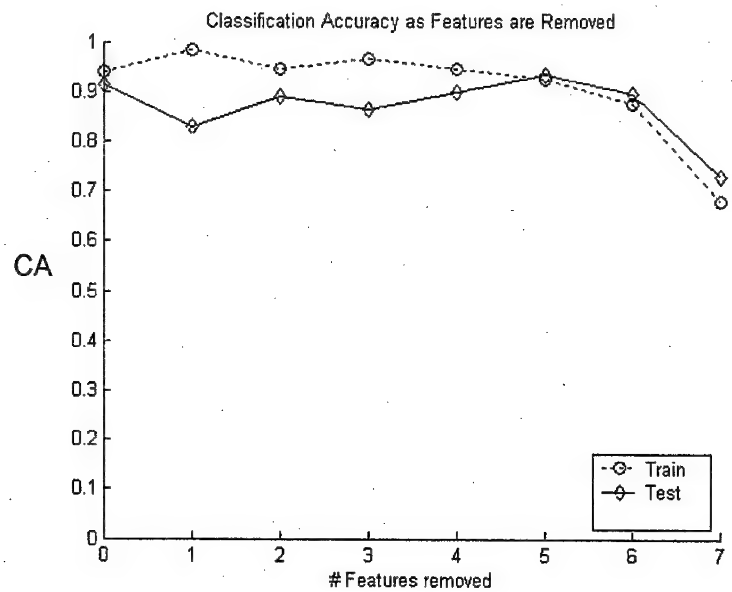

Figure 10. Classification accuracy for training and training-test sets as input features are removed for an Elman RNN and the SNR feature saliency measure.

After performing 20 saliency screening runs, similar results were obtained with both input feature saliency measures. The average number of parsimonious salient features suggested by the SNR method was 2.85 with standard deviation 1.42, while the SBP method suggested 3.15 features with standard deviation 0.99. The SNR method produced a 83.9% mean test set CA with 5.2% standard deviation across the 20 training sets, while the SBP method resulted in a 85.1% mean test set CA with 3.7% standard deviation. The suggested parsimonious features were consistent between the two methods, with both selecting the "blue" features with the two lowest noise levels and "red" feature with the least noise. Both methods also screened out a majority of the 2 known distracter noise features across the 20 replications, with one distracter noise feature included in only 3 of the 40 parsimonious sets. While the SBP did achieve slightly better test set CA and lower variance, the respective SNR values were obtained in an RNN that always included the injected reference noise feature. Thus, both methods should only be compared based on the suggested parsimonious sets, which were equivalent for this limited experiment, or on computational efficiency.

The computational complexity for each saliency method was assessed using the CPU time required to perform the experiment as a surrogate measure for the number of operations required. A dedicated 900 MGHz PC was used to perform all experiments. While each method required unique calculations, both appear to have equivalent computational complexity with an observed difference in mean CPU time less than 2.5%. The mean time for the 20 SNR saliency screening runs was μ_{SNR} = 1719 sec with σ_{SNR} = 219 sec, while the SBP screening mean time was μ_{SBP} = 1680 sec with σ_{SBP} = 212 sec. Performing a two-sample t-test (Wackerly *et al.* 1996) with H_o: $\mu_{SNR} = \mu_{SBP}$ and H_a: $\mu_{SNR} \neq \mu_{SBP}$ statistical evidence is not present supporting H_a. Test statistic T = 0.572 < critical value 2.02, with $\alpha = 0.05$ and associated p-value 0.571.

Differences in computations include the calculation of SNR_i based on summation of weights as in (1), while the SBP method requires calculation of m+1 RNN outputs to compute S_i, the saliency of $i = 1 \ldots m$ input features, at each iteration of the screening algorithm. The

weightspace initialization was also different between methods, leading to differences in the backpropagation learning algorithm. Uniform random first layer weights are required for SNR screening, while the SBP method implemented Nguyen & Widrow (1990) initialization for all weights. The SNR method also adds an additional noise input feature, increasing the weightspace by 8 dimensions (the number of hidden nodes), regardless of the number of features removed. Since the computation time required was statistically equivalent using identical training and test data sets for both methods, network weight optimization, dependent on the stochastic weight initialization, is hypothesized as the primary difference in operations required. If so, the backpropagation gradient descent training algorithm using an approximation of the temporal component, appears to dominate the total number of computations required. A thorough discussion of recurrent networks training algorithms can be found in (Pearlmutter 2001). In summary, differences in computational efficiency do not indicate a preference for weight based SNR or output based SBP saliency screening.

Comparison of Complete and Parsimonious Input Feature Sets

To assess the utility of feature screening to improve RNN classification accuracy, 20 RNNs were trained using all 8 features and the parsimonious set of 3 input features. Classification accuracy is presented by time period in Figures 11-13 and numerically presented in Tables 4-6. In general, these graphs show an upward trend in the mean CA as more observations of a given satellite are observed through time. The observation at time = 0 represents the average of periods 1 through 10. The dashed lines represent a 95% confidence interval for CA across the 20 replications. The training set CA is plotted in Figure 11 and shows minimal classification accuracy differences for both the observed means and variances obtained using the complete and reduced input feature sets. An increased variance associated with wider 95% confidence intervals is observed around time periods 5 and 6 where the "blue" features are not separable for the two true classes.

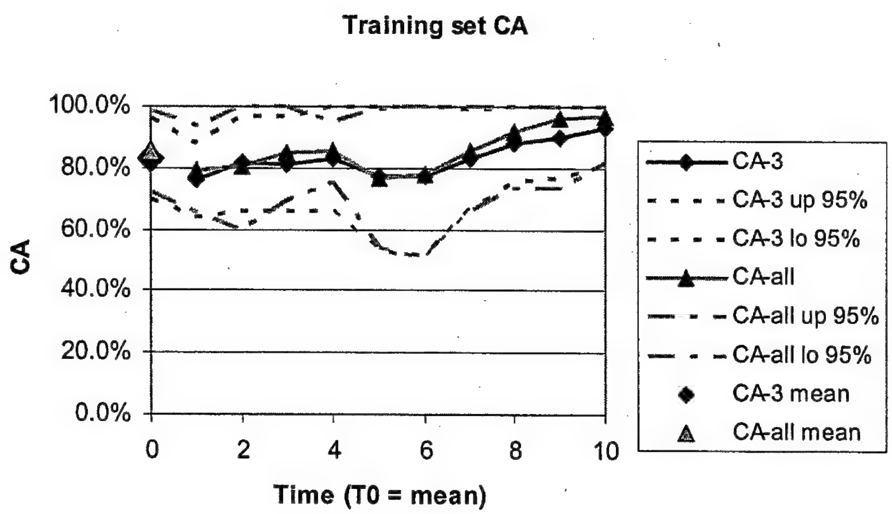

Figure 11. Training set classification accuracy and confidence intervals across 10 time periods using all 8 input features and the parsimonious set of 3 features.

Figure 12 is similar to Figure 11 and shows the training-test set CA and associated 95% confidence intervals. Again, little difference is observed between using the full set of all 8 features and the reduced set of 3 salient features. As expected, the overall training-test set CA is slightly lower than the training set CA. As can be seen in Tables 4 and 5, CA decreases from about 85% down to 80% but only a slight increase in the standard deviation from about 6.5% to 7.6% is observed. Also observed is a "smoothing" effect where the overall CA steadily increases through time and the associated variance does not appear to increase as drastically around periods 5 and 6 in Figure 12 as compared to Figure 11.

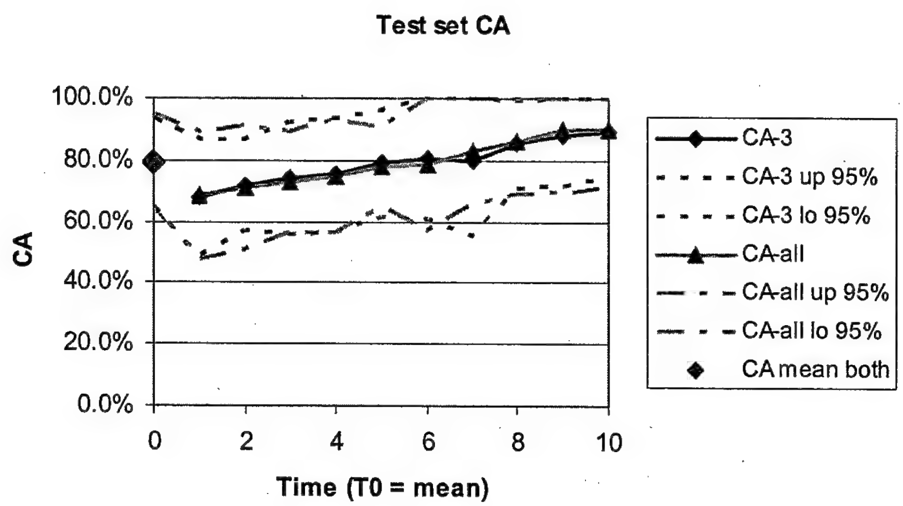

Figure 12. As in Figure 11, classification accuracy and confidence intervals presented for the Training-test set used to stop network weight training.

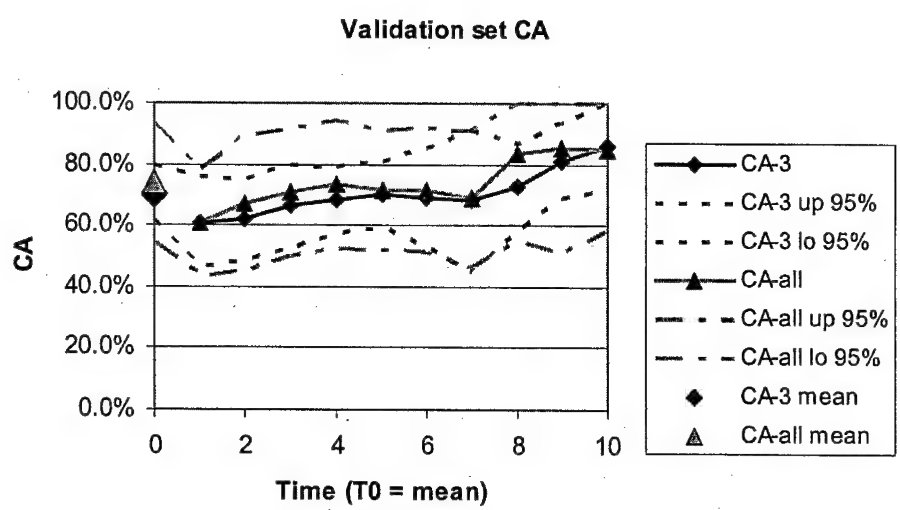

Figure 13. As in Figure 11 & 12, classification accuracy and confidence intervals presented for the Validation set used to assess generalization of the RNN.

Figure 13 contains the CA and 95% confidence intervals for the independent validation set used to measure the RNN's ability to generalize and properly classify new observations. As seen in Figures 11 and 12, the best mean CA is observed in time period 10 when the RNN has been allowed to process all data corresponding to a distinct satellite observation through time. While the average CA using all input features and the reduced set of 3 features are very similar, the 95% confidence interval about the mean CA for the reduced feature set is much narrower. Thus, by using the reduced features lower variance is obtained for the validation set CA at most time periods, inclusive of period 10 where the best CA is obtained.

Table 4. Training set classification accuracy by time period and input variable sets.

All input	Average	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
mean	85.6%	79.5%	80.8%	84.8%	85.3%	76.8%	78.3%	85.5%	92.0%	96.0%	97.0%
stdev	6.5%	7.1%	10.5%	7.9%	5.0%	11.4%	13.5%	9.7%	9.2%	11.3%	7.5%
3 input	**Average**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
mean	83.1%	76.0%	81.5%	81.3%	83.3%	77.3%	77.3%	82.8%	88.3%	90.0%	93.3%
stdev	6.6%	6.0%	7.6%	7.8%	8.6%	12.3%	12.9%	8.2%	6.5%	6.7%	5.9%

Table 5. Training-test set classification accuracy by time period and input variable sets.

All input	Average	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
mean	79.3%	68.8%	71.0%	72.8%	74.8%	78.0%	78.8%	83.3%	86.0%	89.8%	90.0%
stdev	7.9%	10.4%	10.1%	8.2%	9.1%	6.6%	10.7%	8.6%	8.5%	10.1%	9.6%
3 input	**Average**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
mean	79.1%	67.8%	72.0%	74.3%	75.3%	79.0%	80.8%	79.8%	85.3%	88.3%	89.0%
stdev	7.3%	9.5%	7.5%	9.2%	9.2%	8.7%	9.8%	12.2%	7.0%	8.2%	7.2%

Table 6. Validation set classification accuracy by time period and input variable sets. Note: the average standard deviation across 20 replications is significantly less for the reduced feature set, and the largest standard deviations are in periods 8-10 using all candidate input features.

All input	Average	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
mean	73.8%	60.8%	67.0%	70.8%	73.5%	71.5%	71.8%	68.8%	83.3%	85.3%	85.0%
stdev	9.5%	8.6%	11.1%	10.4%	10.4%	9.9%	10.2%	11.3%	14.2%	16.9%	13.3%
3 input	**Average**	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
mean	70.6%	61.0%	62.0%	66.3%	68.5%	70.3%	69.3%	68.3%	72.5%	81.3%	86.3%
stdev	4.5%	7.5%	6.8%	6.9%	5.4%	5.5%	8.0%	11.5%	7.2%	6.3%	7.2%

Table 6 presents the Validation set CA by time period with the first gray block providing the average statistics across all 10 time periods. Table 6 shows use of a reduced feature set lowers the mean CA by about 3%, but results in a desirable reduction in CA standard deviation by more than half (4.5% vs. 9.5%). Also, while use of all input features provides good CA in periods 8-10, the corresponding variance is the largest magnitude for all trained RNNs. Thus, depending on the specific application and decision consequences, a lower mean CA may be preferred with an associated lower variance, such as using the 3 input features in period 9. Finally, a win-win situation occurs if decisions are based on all observations through period 10, where using the reduced feature set has the more desirable mean CA (86.3% vs. 85.0%) and a significantly lower standard deviation (7.2% vs. 13.3%) compared to using all 8 features.

CONCLUSION

While some literature suggests network performance based saliency measures are favorable to weight based saliency measures (Feraud & Clerot 2002; Mak & Blanning 1998) this experiment has shown the SNR and SBP saliency measures perform very similar using an Elman RNN for a specific application. One primary concern of weight based measures is the over-saturation of some weights which provide limited improvement in the overall network (Feraud & Clerot 2002) and (Mak & Blanning 1998). In this experiment and previous efforts using weight based measures (Bauer *et al.* 2000; Greene *et al.* 1997, 2000; Laine *et al.* 2002) networks are initialized with uniform random weights, with network training stopped based on the performance of a training-test set to prevent overtraining of the training set data. Thus, with a proper training algorithm implemented, SNR feature screening based on input layer weights appears consistent to SBP using saliency measures derived from the classification model's output for an Elman RNN.

Both methods converged to the selection of the same parsimonious set of 3 salient features, with the computational efficiency found equivalent both statistically and in practical terms with an observed difference less than 2.5% in mean CPU time required. The parsimonious feature set was compared against use of all candidate input features. By comparing the CA of the training, test, and validation sets, similar CA was obtained using either input feature sets. With validation set CA used as the preferred measure of the RNN's ability to generalize to new data, a significant difference was observed in the CA variance across 20 replications. Use of the reduced feature set led to the desirable reduction in CA standard deviation by over 50%. This limited experiment has demonstrated feasible use of an Elman RNN with a candidate set of input data with significant autocorrelation, crosscorrelation and noise for a classification problem.

This type of feature saliency findings may have substantial benefits for system design or ISR system employment, when reduced feature sets retain or improve an ATR's performance. In these cases, feature selection can provide a means of sensor selection to tradeoff the costs associated with developing, procuring, deploying or simply dedicating a sensor to observe a potential combat target. For the experiment performed, selection of 3 salient input features implies only the sensors collecting that information should be selected for the CID Identity Level Fusion. In addition, the feature selection provided overwhelming evidence to exclude the 2 noise variables which were only retained by 3 of the 40 parsimonious feature sets. Feature selection of spectral data may also help identify optimal bands within the visible and IR spectrum, allowing a multispectral system to be optimally designed that may be less expensive, produces smaller datasets and has a greater SNR ratio for the task-at-hand. Thus, practical benefits may be realized through continued research of feature selection to determine what collected data should be fused for optimal target classification and what associated sensors should be selected. Future research is envisioned with applications of more realistic and demanding data sets including fusion of temporal data sources.

In addition, neural networks may prove to be a useful tool to perform Feature Level, Identity Level or Decision Level 1 Fusion processes when decisions about a single potential target should not be forced. Research by Storm (2003) has demonstrated use of a probabilistic neural network (PNN) as being an effective fusion tool given input data for potential targets with significant correlation between features. Some benefits of using neural models include the ability to fuse data of unknown correlation levels and obtaining continuously valued estimates of class

membership that can be used as a measure of confidence for class membership. Neural models can also be used to fuse input data representative of features, class estimates, or labels all within a one big net (OBN) architecture. Further research shall analyze and explore optimal rules to determine thresholds for the assignment of sensed objects as specific Target types, Non-targets or Unknowns, where more information is required before a confident decision can be made for an Unknown.

ACKNOWLEDGEMENTS

The authors would like to thank Mr. Charles Sadowski from ACC/DRSA for his support and insightful perspectives for this combat ID research. This research is jointly sponsored by ACC and AFOSR.

LIST OF ACRONYMS & SYMBOLS

ACC	Air Combat Command
AFDD	Air Force Doctrine Document
AFOSR	Air Force Office of Scientific Research
ANN	Artificial Neural Network
ATR	Automatic Target Recognition or Recognizer
ATD/R	Automatic Target Detection/Recognition
CA	Classification Accuracy
CID	Combat Identification
COG	Center of Gravity
CPU	Central Processing Unit
DAI	Data In
DAO	Data Out
DEI	Decision In
DEO	Decision Out
DoD	Department of Defense
EOC	Extended Operating Condition
FEI	Feature In
FEO	Feature Out
FLIR	Forward Looking Infrared
HCI	Human Computer Interface
HSI	Hyperspectral Imagery
ID	Identification
I/O	Input/Output
ISR	Intelligence, Surveillance and Reconnaissance
JDL	Joint Directors of Laboratories
MGHz	Megahertz
MLP	Multilayer Perceptron
MSE	Mean Square Error
MSI	Multispectral Imagery
OODA	Observe, Orient, Decide, Act

OBN	One Big Net
PC	Personal Computer
PNN	Probabilistic Neural Network
RNN	Recurrent Neural Network
ROI	Region of Interest
SAR	Synthetic Aperture Radar
SBP	Sensitivity Based Pruning
SNR	Signal-to-Noise Ratio
UAV	Unmanned Aerial Vehicle
UK	United Kingdom
USAF	United States Air Force
α	Confidence level
t	Time period
$f(a)$	Neural network activation function
H_a	Alternative hypothesis
H_o	Null hypothesis
I	Number of input nodes
J	Number of hidden nodes
K	Number of output nodes
M	Number of input features
$MSE(x_{ip})$	MSE of neural network for all p exemplars
$MSE(\overline{x_i})$	MSE of network when an average value is assigned to input feature i
μ	Mean of a statistical distribution
P	Input data dimensionality
p	Number of input exemplars (training data observations)
p-value	Probability of obtaining the test statistic given H_o is true
S_i	Value of the SBP saliency measure for feature i
SNR_i	Value of the SNR saliency measure for feature i
σ	Standard deviation of a statistical distribution
T	Test statistic
$w^1_{i,j}$	Weight from input node i to hidden node j
$w^2_{j,k}$	Weight from hidden node j to output node k
x^1_0	Hidden layer bias term
x^1_j	Output of hidden node j
x^n_0	Input layer bias term
x^n_i	i^{th} input feature of the n^{th} input vector
x_N	Input noise feature
z^n_k	Network output from the k^{th} node for the n^{th} input vector

REFERENCES

Bauer, K.W., Alsing, S.G. and Greene, K.A. "Feature screening using signal to-noise ratios", *Neurocomputing*, Vol 31: 29-44 (Mar 2000).

Bedworth, M.D. "Source Diversity and Feature-Level Fusion," *Proc. of Information Decision and Control 99*, R. Evans *et al.* (eds.), IEEE: 597-602 (1999).

Bedworth, M. and O'Brien, J. "The Omnibus Model: A New Model of Data Fusion?" *IEEE Aerospace and Electronic Systems Magazine*, vol. 15 No. 4: 30-36 (April 2000).

Bishop, C. M. *Neural networks for pattern recognition.* Oxford: Clarendon Press, 1995.

Boyd, J.R. "A Discourse on Winning and Losing," unpublished briefing slides and notes, M-U 43947, Air University Library, Maxwell AFB, AL, 1987.

Calvert, D. and Kremer, S.C. "Networks and Adaptive State Transitions," *A Field Guide to Dynamical Neural Networks*, Kolen, J.F. and Kremer, S.C. (ed.). New York, NY: IEEE Press: 15-25 (2001).

Castellano, G. and Fanelli, A. M. "Variable selection using neural-network models," *Neurocomputing*, Vol 31: 1-13, (Mar 2000).

Chitroub, S., Houacine, A. and Sansal, B. "Statistical characterization and modeling of SAR images," *Signal Processing,* Vol 82: 69-92 (2002).

Costantini, M., Farina, A. and Zirilli, F. "The fusion of different resolution SAR images," *Proceedings of the IEEE,* Vol 85, No. 1: 139-146 (1997).

Dasarathy, B.V. "Sensor Fusion Potential Exploitation—Innovative Architectures and Illustrative Applications," *Proceedings of the IEEE,* Vol 85, No 1: 24-38 (January 1997).

Department of the Air Force. *Air Warfare*. AFDD2-1. Washington, D.C.: HQ USAF, January 2000.

Department of the Air Force. *Intelligence, Surveillance, and Reconnaissance Operations*. AFDD2-5.2. Washington, D.C.: HQ USAF, April 1999.

Department of the Air Force. *USAF Intelligence Targeting Guide*. AFP 14-210. Washington, D.C: HQ USAF, February 1998.

Dudgeon, D. E. *ATR performance modeling and estimation.* Tech. Rep. 1051, Lincoln Laboratories, Massachusetts Institute of Technology, Lexington, MA: 1998.

Elman, J. L. "Finding Structure in Time," *Cognitive Science*, Vol 14: 179-211 (1990).

Feraud, R. and Clerot, F. "A methodology to explain neural network classification," *Neural Networks*, Vol 15: 237-246 (2002).

Gat, N., Subramanian, S., Barhen, J. and Toomarian, N. "Spectral imaging applications: remote sensing, environmental monitoring, medicine, military operations, factory automation and manufacturing." Proceedings of SPIE Vol. 2962: 63-77 (1997).

Giles, C.L. and Omlin, C., "Representation of Discrete States," *A Field Guide to Dynamical Neural Networks*, Kolen, J.F. and Kremer, S.C. (ed.). New York, NY: IEEE Press, 2001, pp. 83-102.

Greene, K.A., *Feature Saliency In Artificial Neural Networks With Application To Modeling Workload,* Dissertation, Air Force Institute of Technology, WPAFB, OH, 1998.

Greene K.A., Bauer, K.W., Kabrisky, M., Rogers, S.K. and Wilson, G.F. "Estimating pilot workload using Elman recurrent neural networks: a preliminary investigation," *Intelligent Eng. Sys. through ANNs*, Vol 7, Proc. of the ANNIE Int'l. Conf.: 703-708 (Nov 1997).

Greene, K.A., K Bauer, K.W., Wilson, G.F., Russell, C.A., Rogers, S.K. and Kabrisky, M. "Selection of psychophysiological features for classifying air traffic controller workload in neural networks," *Smart Eng. Sys. Design*, Vol 2: 315-330 (2000).

Hall, D.L. and Llinas, J. (eds.). *Handbook of Multisensor Data Fusion.* New York: CRC Press, 2001.

Hall, D.L. and Llinas, J. "An Introduction to Multisensor Data Fusion," *Proceedings of the IEEE*, Vol 85, No 1: 6-23 January 1997.

Haspert, K.J. "Optimum ID Sensor Fusion for Multiple Target Types." Institute for Defense Analysis (IDA) Document D-2451, Virginia: March 2000.

Hornik, K., Stinchcombe, M. and White, H. "Multilayer feedforward networks are universal approximators," *Neural Networks,* Vol 2: 359-368 (1989).

Hornik, K., Stinchcombe, M. and White, H. "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks," *Neural Networks,* Vol 3: 551-560 (1990).

Jackson, Q. and Landgrebe, D. "An adaptive classifier design for high-dimensional data analysis with a limited training data set," *IEEE Trans. on Geoscience and Remote Sensing,* Vol 39, No 12: 2664-2679 (December 2001).

Jimenez, L.O. and Landgrebe D.A. "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Trans. on Sys., Man and Cybernetics C,* Vol. 28, No. 1: 39-54 (1998).

Kremer, S.C. "On the Computational Power of Elman-Style Recurrent Networks." *IEEE Transactions on Neural Networks*, Vol 6, No. 4: 1000-1004 (1995).

Kwak, N. and Choi, C.H. "Input feature selection for classification problems," *IEEE Trans. on Neural Networks*, Vol. 13, No 1: 143-159 (2002).

Laine, T.I. and Bauer, K.W. "Feature Selection Assessment and Comparison using Two Saliency Measures in an Elman Recurrent Neural Network," *Proc. Int'l Joint Conf. on Neural Networks,* IEEE Press, 2807-2812 (2003).

Laine, T.I., Bauer, K.W., Lanning J.W., Russel, C.A. and Wilson G.F. "Selection of input features across subjects for classifying crewmember workload using artificial neural networks," *IEEE Trans. on Sys., Man and Cybernetics A,* Vol. 32, No. 6: 691-704 (2002).

Landgrebe, D. "Hyperspectral image data analysis as a high dimensional signal processing problem," *IEEE Signal Processing Magazine*, Vol. 19, No. 1: 17-28 (January 2002).

Lazzerini, B. and Marcelloni, F. "Feature selection based on similarity," *Electronics Letters,* Vol 38, No. 3: 121-122 (2002).

Lee, J., Hoppel, K.W., Mango, S.A. and Miller, A.R. "Intensity and phase statistics of multilook polarimetic and interferometric SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing,* Vol 32, No 5: 1017-1028 (1994).

Looney, Carl G. *Pattern Recognition using Neural Networks.* New York: Oxford University Press, 1997.

Mak, B. and Blanning, R.W. "An empirical measure of element contribution in neural networks," *IEEE Trans. on Sys., Man, and Cybernetics: Part C,* Vol 28: 561-564 (Nov 1998).

Moody, J. "Forecasting the economy with neural nets: a survey of challenges and solutions," *Neural Networks: tricks of the trade*, Orr, G.B and Muller, K.R. (ed.). New York, NY: Springer: 348-371, 1998.

Nasr, H. N. "Fundamentals of Automatic Target Recognition," *Short Course Notes,* SPIE SC 158 (April 2003).

Nguyen, D. and Widrow, B. "Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights," *Proc. Int'l Joint Conf. on Neural Networks*, Vol 3: 21-26 (1990).

Pearlmutter, B.P. "Gradient calculations for dynamic recurrent neural networks," *A Field Guide to Dynamical Neural Networks*, Kolen, J. F. and Kremer, S. C. (ed.). Piscataway, NJ: IEEE Press: 179-206, (April 2002).

Piramuthu, S. "Feature selection for financial credit-risk evaluation decisions," *INFORMS J. on Computing*, Vol 11, No 3: 258-266 (Summer 1999).

Ralston, J.M. "Bayesian Sensor Fusion Minimum-Cost I.D. Declaration." Institute for Defense Analysis (IDA) Paper P-3441, Virginia: June 1999.

Ross, T., Bradley, J., Hudson, L. and O'Conner, M. "SAR ATR – So what's the problem? – An MSTAR perspective." Proceedings of SPIE Vol. 3721: 662-672 (April 1999).

Ruck, D.W., Rogers, S.K. and Kabrisky, M. "Feature selection using a multilayer perceptron," *J. of Neural Network Computing*: 40-48 (Fall 1990).

Schroeder, J. "Automatic target detection and recognition using synthetic aperture radar imagery." *Proc. of the 2002 Workshop on the Applications of Radio Science (WARS02) Conf.*, Wilkinson, *et al.* (eds.), National Committee for Radio Science, Australia, 2002.

Steinberg, A.N., Bowman, C.L., and White, F.E. "Revisions to the JDL Data Fusion Model." *Proceedings of AeroSense 1999*, SPIE Vol. 3719: 430-441 (1999).

Storm, S. *An Investigation of the Effects of Correlation in Sensor Fusion.* Thesis, Air Force Institute of Technology, WPAFB, OH, 2003.

Tarr, G.L. *Multilayered Feedforward Neural Networks For Image Segmentation*, Dissertation, Air Force Institute of Technology, WPAFB, OH, 1991.

Utans, J., Moody, J., Rehfuss, S. and Siegelmann, H., "Input variable selection for neural networks: application to predicting the U.S. buisiness cycle," *Proc. of the IEEE/IAFE 1995 CIFE.* New York, NY: IEEE Press, 118-122 (1995).

Verikas, A. and Bacauskiene, M. "Feature selection with neural networks," *Pattern Rec. Letters,* Vol 23: 1323-1335 (2002).

Waltz, E. and Llinas, J. *Multisensor Data Fusion.* Boston, MA: Artech House, 1990.

Wackerly, D. D., Mendenhall, W. and Scheafer, R. L. *Mathematical Statistics with Applications.* Belmont, CA: Duxbury Press, 1996.

Young, M., Dimalanta, A., Konno, D., Peli, T. and Zachery, K. N. "Shade Feature Level Fusion Results," *MSS Nat'l Symp. on Sensor and Data Fusion*, Titan System Corp., June 2001.

Zhang, H. and Sun, G. "Feature selection using tabu search method," *Pattern Recognition*, Vol 35: 701-711 (2002).